DOCUMENT RESUME

ED 185 991 IR 008 321

AUTHOR Clement, Hope E. A.
TITLE The Automated Authority Files at the National Library of Canada.
PUB DATE Aug 79
NOTE 29p.; Paper presented at the Conference of the International Federation of Library Associations and Institutions (45th,, Copenhagen, Denmark, August 27-September 1, 1979). For related documents, see IR 008 333-344. Legibility varies.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cataloging; Classification; Foreign Countries; *Library Automation; *Library Catalogs; Library Equipment; *Library Networks; Library Technical Processes; *Online Systems; Telecommunications
IDENTIFIERS *Authority Control (Information); Canada; *Dortmunder Bibliothekssystem

ABSTRACT

This paper discusses the current resurgence of interest in authority files and describes the automated authority files of the National Library of Canada. Factors contributing to this interest in such files include cooperative programs between libraries, the development of international standards, and the impact of computer technology and telecommunications. An automated authority file can result in great savings in time and money, as well as maintaining file integrity while allowing flexibility for local usage. The National Library of Canada, which has always had manual authority files, currently uses a batch mode authority system. A pilot project online authority system, DOBIS (Dortmunder Bibliothekssystem), is now operating at the National Library and the Canada Institute of Scientific and Technical Information for the production of a union list of serials and for reference activities. Completion of the pilot program is expected in September of 1979 with expanded use of the system planned for the year's end. The DOBIS system has been designed to provide decentralized access to all catalogs and files; it is a fully integrated system incorporating cataloging, catalog search, circulation, acquisition, and information retrieval functions. (RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185991

IFLA COUNCIL 1979

Division: Management and Technology

Section: Mechanization

17 INEC IE
17 CAT 1E

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Automated Authority File at the National Library of Canada

by

Hope E.A. CLEMENT
Associate National Librarian
National Library of Canada
Ottawa, Ontario

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carol Kucen Henry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR008321

1. Why Authority Files?

In bibliographic and cataloguing circles, authority files are becoming one of the most frequent topics of discussion, especially in connection with automated bibliographic networks. Why this resurgence of interest in what is basically an old concept? Libraries have maintained manual authority files for years to ensure consistency in cataloguing practice, using master official catalogues or separate authority files of headings, or a combination of the two.

I believe a number of factors are contributing to this recent interest in automated authority files. First, libraries are realizing that they cannot continue to exist as separate entities and still succeed in supplying their users with the world's literature. The sheer volume of publishing and, in many areas of the world, the decrease in real purchasing power of libraries resulting from inflation and currency fluctuations, are leading to co-operation between libraries as never before. In this spirit of cooperation, the UBC and UAP programmes have been established. To succeed, international bibliographic exchange requires a high degree of standardization. Authority files, created according to accepted standards, are one means of ensuring that records from many sources can be successfully matched against one another and added to data bases.

Increased cooperation both exploits and depends upon new technology, particularly in the areas of telecommunications and automation. Successful cooperation requires frequent and rapid communication between the various parts of a network, as is possible with on-line computer systems, which permit immediate interaction among users and creators of data bases. In order to preserve data base integrity, it is essential that standards be agreed upon and enforced, and that the role of various nodes in a network be clear, particularly regarding the right to change, add to or delete from a data base. This has led to the concept of authentication centres which are assigned the responsibility of verifying specific parts of a data base, particularly access points.

Although automation has greatly expanded the possibilities of library catalogues by adding the ability to manipulate and search a bibliographic data base and to produce from it a variety of products customized to the needs of various users, it has also introduced a new element of rigidity into cataloguing. By itself, a computer cannot reconcile variant names for the same entity when matching bibliographic records. Despite the use of compatible MARC formats and cataloguing rules, and of complex coding, this fact has made the need for further standardization urgent, a standardization which can accurately ensure the collocation function in a large file, but which permits flexibility in local usage and in cataloguing products. Duplication of records in large data files is proving costly in searching time as well as frustrating for users. Authority files are seen as a solution.

The rate of change in the modern world also has significant effects on library catalogues. The cost of implementing changes in large data bases is high, both in reprogramming to handle format updates and in changes to the

data itself. It is essential that there be economical ways of updating data bases as records and terminology change. For instance, verbal subject headings soon become obsolete and new terms must be added constantly to keep up with new areas of development and interest. Names, particularly corporate names such as those of government bodies, also have a high rate of change. New bibliographic standards, for example a new cataloguing code like AACR2, will affect the form of a considerable number of headings in a data base. The need to adapt bibliographic records received from other countries to national usage and language is also an important factor. An automated authority file can be the means of carrying out such changes economically, while maintaining data base integrity and still allowing national and local flexibility. As well, savings can be made in eliminating data redundancy in headings that are used over and over again. A short study of headings being used for current National Library cataloguing has shown that 57.5% of the headings required were already recorded in the authority file, with a rate of 80% recurrence for names used in government publications. Thus, it is cost effective to avoid the repetitious input and storage of these names.

The National Library of Canada has always had manual authority files, providing considerable detail on Canadian authors. In the early days of Canadiana, the national bibliography, before automation, we used files of numbered sets of headings and their related cross-references for government corporate names, inserting these headings in the bibliographic records during the layout of camera-ready copy, thus saving considerable typing and proofing time. In Canada, with our federal system and offical bilingualism, we have two official names, English and French, for federal government bodies, and, in addition, we have to cope with ten provinces and their government departments. So we have always been very conscious of the repetitious and complex nature of corporate headings. This fact was kept in mind in determining national and bilingual bibliographic standards for Canada.

2. Background of the National Library Authority Subsystem

During the early 1970's, when the Canadiana/Cataloguing System, an automated batch system, was being developed and implemented at the National Library, the special requirements of a bilingual country were taken into account. The system was developed to handle both in-house cataloguing and the production of the national bibliography, Canadiana, and its related products and services. Bilingual cataloguing was implemented: English works are catalogued in English, French works in French, and bilingual publications have two cataloguing records, one in English, the other in French. All headings are created in both official languages where appropriate, in the form prescribed by the English and French versions of the Anglo-American Cataloguing Rules. A necessary requirement, in order to ensure that all works belonging under one heading are located together regardless of language, is the linking of equivalent language headings, both name and subject, through an authority file subsystem. The Canadian MARC format thus provides that equivalent name headings in the other official language can be stored in the 9XX fields of bibliographic records; equivalent subject headings are stored in the 6XX fields. The language of each is identified by second indicator values.

The National Library's automated authority subsystem was designed to complement the cataloguing system which now comprises two subsystems, bibliographic and authority. The first objective was to control a variety of bilingual authority files, covering both name and subject headings, and ensure effective links between equivalences in both official languages. The second objective was to interface with and control the Library's machine-readable bibliographic file. Together, these objectives would ensure effective links between languages and the generation of a number of different bibliographic products to meet a variety of needs, both internal and for national and international distribution.

3. The Batch-Mode Authority Subsystem Currently in Operation

In order to meet the stated objectives, the authority subsystem was designed with five principal capabilities and characteristics:

1. A central file of authority data was created and maintained,

2. Searching techniques were devised, using the normalized key technique, to ensure the uniqueness of each heading added to the file, thus preserving the essential collocation characteristic of the subsystem,

3. Relationships between headings, language equivalences and cross-references were identified and maintained in the file and could be automatically generated for updating or revising headings,

4. Authority data were linked to and could be automatically combined with the appropriate bibliographic records to produce any of the variety of products possible from the subsystem, for both in-house use and outside distribution.

5. Various authorities products could be generated directly from the authority file data, including a nationally distributed authority service on COM,

6. Flexibility was provided to meet the needs of a multi-library or network envirionment.

In the authority subsystem, in the internal authority format,[1] each heading or reference is stored as a separate record. Each record has a unique

1 An authority record in the National Library's internal MARC format consists of a 90-character leader, a variable length record directory, variable fields, and linkage fields. Each variable field contains subfield codes at the beginning followed first by a control subfield "w", and then by authority data. Apart from subfield "w", which is unique to the authority record format and the linkage fields, this structure is identical to that employed for bibliographic records in the internal National Library formats. The directory entries consist of a three character tag, and indication of the specific occurrence of that tag, an error identifier, the length of the variable field, and the starting character position of the variable field.

authority control number, which is simply an identification device. This number is used to link the authority record to the bibliographic records in which the heading occurs as an access point. Thus, in the bibliographic record, an authority record control number replaces the actual heading. The control number is also used to connect related authority records, to cover "see", "see also", "see from" references, and so on. Each heading, whether an accepted heading or a variant, along with its equivalent in the other official language, the relevant notes, filing information, and control data, comprises a single authority record. The actual use of each of these headings is specified in the control fields of the authority record, e.g. use as an accepted heading or as a variant form of name used only as a reference to the accepted heading.

The linkages with bibliographic records are achieved by storing the appropriate control number within the entry field of the bibliographic record. The linkages with other related authority records and cross-references are made by storing the authority control numbers of these linked records within special linkage fields, the 9XX fields, in each authority record. In addition to the normal kinds of relationships between authorities just mentioned, the 9XX fields also permit storage of "hierarchy to" and "hierarchy from" references for names. Thus, linkages are used to supply cross-references to the general level of a corporate name as well as to headings involving a subdivision of the name, for instance, the reference Toronto University - needed for the general heading, University of Toronto can be supplied also for the more specific subdivision University of Toronto. Department of Chemistry. The subsystem normally stores the English and French equivalences of a name or subject heading in the same record, with the same control number.

In all cases, a heading or cross-reference occurs only once in the entire authority file. A "normalized key" is generated for each heading or reference added to the file, and this is used as a unique matching device to discover duplicate headings and references. The algorithm used to generate this key is based on the one developed for the Authority System of the New York Public Library.[2] All incoming headings are machine-matched against existing headings and an incoming heading and its resultant normalized key will be added to the authority file only if the match based on the normalized key fails to meet an identical normalized key within the authority file.

The features of the system as outlined above have many advantages due, in main, to the fact that bibliographic files and authority files are effectively linked. The concept of one unique record for each heading or reference reduces the amount of redundant data in the authority file, and reduces the possibility of erroneous duplication of headings. The use of

2 The normalized key algorithm is created by first eliminating diacritics, extraneous spaces, punctuation and subfield coding. All characters in the heading are then converted to their equivalent upper case EBCDIC representation, with any nonblank characters past the 80th input to a hashing algorithm, which takes into account such elements as language.

authority control numbers in bibliographic records in the place of a name and its network of references, represents a great saving in time required to code, input and proofread data, and substantially reduces the size of bibliographic records. Further, and perhaps most important, a revision to a heading is made only once, in the appropriate authority record, and then is reflected in all products which are generated from the bibliographic records which relate to that particular authority record. This feature will be of particular benefit when the National Library adopts the second edition of the Anglo-American Cataloguing Rules, in 1981, requiring the revision of a substantial percentage of existing headings now in the file. The capabilities of the system will enable the National Library to switch to the new rules with comparative ease, permitting the update of existing bibliographic records in the machine-readable file with a minimum of difficulty and staff-time.

I have mentioned that one of the advantages of the National Library's authority subsystem is that it controls headings for both in-house use and for national bibliographic products to be distributed nationally and internationally. Normally, authority files are library-specific, that is, they record a library's established form and use of particular headings. In order to build a file which would be potentially useful both in-house and to other libraries, the National Library adopted a "neutral file" principal for establishing headings. The phrase "neutral file" is used to indicate that authority records are created according to a national standard and do not necessarily reflect the usage of individual libraries in a network. At present, the use of the heading and equivalence data in each authority record can be separately specified for up to six different users, libraries, or files. This is accomplished by storing codes indicating local use, in specialized fixed fields within the authority control record. Using these library specific control features, it is possible to indicate that a certain form of name heading is acceptable in one user's file, but is only a cross-reference in another user's file. This control capability, for example, can permit slightly different name and subject authority usage in the National Library's own in-house catalogues from that used in national bibliographic products to be distributed outside, when desirable, to reflect the difference between a catalogue and a bibliography. The Canadiana records are those of the neutral file. They are the Canadian standard records, created according to national and international standards, without reference to the practices or files of a particular library. The standards presently used by the National Library in the creation of headings for the neutral authority file are as follows:

1. Anglo-American Cataloguing Rules (Revised) (English and French versions) for personal and corporate names, and headings for series, title and author/titles. Use of the second edition of AACR2 will begin in 1981.

2. Library of Congress Subject Headings (most recent edition and supplements) for English language non-Canadian subject headings.

3. Canadian Subject Headings (published by the National Library, 1979) for English language Canadian subject headings.

4. Répertoire de vedettes-matières (8th edition, published by the Université Laval, Québec) and supplements for French language subject headings.

At present, the authority subsystem operates in a batch mode, and still interfaces with authority files in card form to some degree. When a new Canadian title is catalogued, two types of record, the bibliographic record and the authority record(s) must be created. If the appropriate authority record already exists in the authority file, then its control number is entered on the bibliographic worksheet. If not, then the authority generation process must be initiated. An authorities worksheet is used for this purpose, with temporary slips attached which can be filed immediately in the manual authority file. This file will be maintained until an on-line authority system is operational. Both the English and French headings are recorded in the temporary area of the worksheet and two temporary slips are filed, one for each language version of the heading.

A linkage field is added to the record for the accepted heading to allow the records for variant forms to be linked to it, and to specify the relationships between the headings. The reverse linkages are machine-generated. The actual content that must be transcribed to create the record for the variant heading is limited because a considerable amount of data is supplied by the record creation program. Thus, an authority record for a variant heading may be created by encoding only two or three fields: control number (001), English heading (210), and French heading (110), if the heading exists in both of Canada's official languages.

Until recently, proofsheets were generated following transcription of the data on the authorities worksheet into machine-readable form and submission to the authority record creation programs. At present, data entry in both the bibliographic and authority subsystems is by means of a IV Phase remote key to disk system, which uses cathode ray terminals controlled by a 192K programmable processor. Fully interactive data entry, edit, and correction of records is supported by IV Phase. The output tape from the IV Phase is then used to update the master files which reside on an IBM 370. The cycling of proof-sheets, so time-consuming in batch systems, is gradually disappearing as more and more records can be proofread, corrected and verified directly on the screens.

The National Library has built a data base file of MARC records as a basis for bibliographic activities. An early step in the development of the data base and its authority file was the SPLIT project, carried out in 1977. This project aimed at using machine-readable bibliographic records already created for Canadiana, in the period 1973-76 before the authority subsystem was developed, to generate skeletal authority records for names, uniform titles and author/title headings. The project also endeavoured to reduce the amount of redundant data in the bibliographic file by replacing the access point strings in the bibliographic records with the authority control numbers of

the skeletal records generated in the SPLIT. This project, which was only a partial success, clearly illustrated a number of the problems which libraries face when undertaking retrospective conversion of files, and in this respect, was a more useful experience. The number of headings generated was overpowering, 61,000 in all, consisting mostly of personal names, many of which would never be used again. However, 15,000 government names, the most frequently used corporate names, were successfully matched against accepted forms, and the bibliographic records associated with these thus automatically updated. Updating procedures of the batch subsystem placed a burden on cataloguers, coders and keyboarders following SPLIT, indeed, upgrading of records took longer than original creation would have. This was true of the batch subsystem, on-line updating would have been much more efficient. The lessons learned in the SPLIT project, concerning the difficulties of retrospective imposition of rigorous control over an existing data file, problems of upgrading records, and the necessity of extensive testing and sampling before such a project is undertaken have been of great value in preparing for the data base load for the on-line system presently being implemented.

These authority subsystem capabilities have contributed to better control, versatility and greater data integrity in both the text and indexes of the printed Canadiana, the national bibliography, and have made possible a range of tape and microfiche services derived from the data base as well as a number of cooperative projects.

The products and services using the authority subsystem are as follows:

1. Bibliographic products

   Bibliographic records from the Canadiana data base are produced and distributed in three forms, printed, microfiche and machine-readable. The combined bibliographic and authority subsystems are used to produce records for the printed national bibliography, Canadiana, issued monthly and cumulated annually; Canadiana records also appear in a COM (Computer Output Microform) version, issued monthly but available more quickly than the printed version and called Canadiana Microfiche; records are also distributed on the weekly CAN/MARC tapes and, as well, specific records for particular items can be ordered either in machine-readable form or as unit catalogue cards.

2. Services providing authority records only are:

   CAN/MARC: Authorities - A quarterly COM service, with bi-weekly supplements, designed to provide libraries with a current list of verified name headings of Canadian origin. Each quarterly issue represents a complete regeneration of the base file, with the supplements providing the latest information on new entries and revised headings.

   Répertoire de vedettes-matière: Supplément - A quarterly cumulating COM service, containing French language topical subject headings and related references created since the publication of the 8th edition of the Répertoire. Indexes, both French-English and English-French, are

included. This is a cooperative project with the Université Laval in Québec.

Beginning in January 1980, we intend to issue authority records in machine-readable form on tape. In order to do this, we are developing a MARC authority communication format which differs somewhat from our internal format. In our internal format, each heading, whether the national standard (accepted) or a variant (reference) is issued as a separate record with its own control number together with its equivalent in the other official language. This enables libraries in a network environment to choose whether to use a heading or one of the variants. In the communication format, a heading and its references will be grouped into a single record with one control number. Two separate records will exist for English and French equivalent headings. The bibliographic records on CAN/MARC tapes will contain the authority control numbers of the access points as well as the access points themselves. This linkage between the authority and the bibliographic records will allow automatic updating of the access points in the latter.

The products, services and capabilities of the authority subsystem have made participation in a number of special cooperative projects both possible and effective. These include the production and update of authoritative lists of Canadian subject headings in both French and English; agreements between the National Library and the Library of Congress on corporate names to reduce conflict in data provided to North American libraries for Canadian imprints; participation in the CONSER project, and in the Canadian Cataloguing-in-Publication (CIP) program.

The CONSER project, established in 1974, is a cooperative effort among 14 American and Canadian libraries to build a serials data base. It was one of the early examples of the practical application of the concept of the "authentication centre". In CONSER, two of the participants, the Library of Congress and the National Library of Canada acted as authentication centres. While all 14 participants contributed records to the CONSER file, only these two institutions could "authenticate" or verify records as conforming to the project standard. The CONSER file is also an early example of a data base with recognized differences in the levels of records - e.g., authenticated, unverified, etc.

Earlier this year, the National Library published a 2nd edition of the authoritative list for Canadian subject headings in English, Canadian Subject Headings. While this edition was produced using manual means, future supplements and editions of the list will be produced using the automated subsystem, deriving its contents from the subject authority files.

Work is continuing on a cooperative undertaking between the National Library and the Université Laval to produce a new edition of a similar authoritative list of subject headings in French. Presently in its 8th edition, work is progressing on the preparation of a 9th edition of Répertoire de-vedettes-matière. Laval acts as the authentication centre for French language headings complementing the Library of Congress English language list, and

creates the basic worksheet data for each heading. The National Library inputs the records to the authority subsystem. By May 1979, approximately 20,500 headings and 7,000 cross-references had been entered on the National Library's subsystem. Approximately 45,000 headings from the 8th edition of Répertoire de vedettes-matière remain to be converted.

The National Library also has an agreement with the Library of Congress regarding corporate name headings. Under this agreement, reached in 1975, the Library of Congress accepts for its own use and for production of its external cataloguing products, name headings established by the National Library for Canadian corporate bodies. The Library of Congress receives the CAN/MARC: Authorities fiche for names, and when a Canadian name which the Library of Congress wishes to create is not found in the COM list, the National Library is contacted, by the use of telefacsimile transmission, for establishment of the name. The National Library authenticates the name, transmits a reply to the Library of Congress and enters the heading in the authority file. Even though they may not yet be used by the National Library itself, all such authenticated names appear on the COM fiche authorities service for the benefit of other users. This arrangement ensures that Canadian usage is maintained in records disseminated by the Library of Congress and that there are no discrepancies between Canadian and American data.

The Canadian Cataloguing-in-Publication (CIP) program was established by the National Library in a decentralized mode in 1976. While the National Library acts as the authentication centre for all data produced by the program, and distributes the products, the actual cataloguing is done on contract by agent libraries in a number of cities across the country close to the major publishing centres. Headings created by CIP agent libraries are verified by the National Library, entered on the subsystem, and made available for external distribution in the CAN/MARC: Authorities service. Until last year, the Canadian CIP program was limited to trade publications. More recently, selected provincial government publications have been added, and there is a possibility that, in future, provincial agencies could become the authentication centres for name headings related to their own provincial jurisdictions, just as the Université Laval has become for French language subject headings. The implementation of an on-line system at the National Library will accelerate and facilitate the operation of such a decentralized CIP program by allowing agent libraries and provincial authentication centres to enter data directly from remote terminals and have immediate access to complete, up-to-date authority files.

Some statistics on the authority subsystem may be of interest. In mid-May 1979, the 100,000th verified record was added to the file, which began in 1975. Some 36,000 personal names and 25,000 corporate names are now represented in the file, together with about 2,000 uniform titles and 20,000 topical subject headings. Cross-reference records of various kinds total 28,000. About 3% of the personal name headings have forms in both English and French, compared to 25% for corporate names (38% for government names). Almost 73% of the subject headings are in both English and French.

About 40,000 records are now added to the authority file each year, compared to about 28,000 bibliographic records. As a result of the SPLIT project, over 135,000 bibliographic records created for Canadiana, since 1973, are now machine-linked to the authority records and their access points may thus be automatically updated.

4. Work in Progress on the Development of the On-line Authority Capabilities of the DOBIS System

The development of the authority subsystem to date can be described as the first step in the introduction of the data base management philosophy to National Library processing. The application of this philosophy will be greatly broadened in the next few years as the implementation of the DOBIS Dortmunder Bibliothekssystem) at the National Library moves through the final stages of pilot projects and into an operational mode.

The main stimulus for the development of an on-line data base management system at the National Library came in the early 1970's from the need to automate our huge manual union catalogue. In 1972, the Canadian Union Catalogue located at the National Library, held 12 million cards, representing about 4 million titles and was expanding at the rate of over 1,500,000 accession reports a year from up to 300 Canadian libraries. Lack of physical space for file expansion, lack of access to the file other than by main entry, lack of sufficient staff to keep up with filing, editing and searching, all indicated that alternative methods of maintaining such a store of data should be investigated. The solution proposed to these problems was an on-line system.

Evaluations were carried out on a number of existing library systems to see if any met the overall requirements of the National Library: a system able to handle all the usual in-house operations of a library, able to satisfy the data base requirements of the proposed union catalogue, to support a shared cataloguing program amongst the National Library and other federal libraries, to generate the national bibliographic products already produced or planned for distribution, and to operate as a central or federal node in a national bibliographic network. In 1975, the DOBIS system (Dortmunder Bibliothekssystem) was selected as the existing system most nearly meeting these requirements. Since that time, the system has undergone a rigorous and extensive evaluation, and major modifications have been completed to make it MARC-compatible and to introduce the English/French bilingual capability necessary in Canada as well as the ability to operate in a network mode.

At present, DOBIS is operating in a pilot project mode at the National Library, and a related federal institution, the Canada Institute for Scientific and Technical Information (CISTI) at the National Research Council. As of April 1979, 384,000 records, including LC, CAN/MARC and CONSER source files, have been mounted on the system. DOBIS is being used at the National Library for input and production of a union list of serials in the humanities and social sciences, for reference activities and for limited searching to find National Library locations for titles requested on

interlibrary loan. At CISTI, a pilot project in cataloguing technical reports is being carried out. Completion of the pilot project phase of the DOBIS project is scheduled for September 1979, with expanded use of the existing modules - catalogue , catalogue search, planned for later in the year. The system includes an authority structure as part of its overall structure, and full implementation will continue to facilitate and improve the authority products and services provided by the National Library, as well as encourage the development of more decentralized networking and cooperative activities. To date, the National Libray has been an active participant in cooperative projects which have explored in a practical way some aspects of networking, in particular, the technical and managerial reuirements for sharing responsibility for the creation of national source records, and DOBIS will be the major vehicle for future ventures in this area. At present, both DOBIS and National Library batch systems are running on a dedicated IBM 370/148 2MB computer, located at the National Research Council. This will soon be replaced by an IBM 3031 computer, which is 2.5 times faster than the current central processor.

The DOBIS system has been designed to provide decentralized access to all library catalogues and files, and to permit one time entry of data. It is a fully integrated system, incorporating the cataloguing, catalogue search, circulation, acquisitions, serials control, and information retrieval functions. To date, the cataloguing and catalogue search modules have been adapted for Canadian federal government use.

Files in DOBIS are organized to support this integrated system concept. Only one copy of a bibliographic record is kept in the system's files, and that one copy is used for all purposes. These bibliographic records are kept in a "document file", and each is assigned a unique DOBIS document number by the system. Similar to the batch system, the access points most commonly used for searching - names, titles, subject headings, etc., are detached from the bibliographic record and placed in separate index files, retaining links to the bibliographic record they come from. Thus, the DOBIS document file contains no complete unit records. When a user asks the system to display a complete bibliographic record, the system automatically assembles the various components of the record from the document and index files and then displays the resulting combination.

This arrangement has similar advantages in DOBIS to those in the batch system: files are consistent, unnecessary duplication is eliminated, and updates and changes are made easily. When a change is required in an authority heading, it is made once to the entry in the access point file, and it is then reflected in all the bibliographic records attached to that entry. This feature greatly facilitates the changing, updating and upgrading of headings. It will be particularly useful when the National Library adopts the new cataloguing rules in 1981, permitting the revision of many bibliographic records through changes in access point files.

Most of the bibliographic and access point data in DOBIS are stored in the main set of shared files, called the system-level files. These files contain data shared by all libraries using DOBIS and are in effect the "neutral"

records. They are maintained according to agreed-upon policies and procedures designed to make the system-level files useful for all library requirements, thereby avoiding the costly duplication of files for each library using the system. However, in some cases, local library needs may not be met by system-level data, and in these cases the local library can establish separate local document and access point files, organized in the same manner as the system-level files. Bibliographic data in local and system DOBIS document files is linked by a sharing of the same DOBIS document number by the system and local records. The user enters a system-level record before entering the variant data for a related local record. Thus the local file contains only the variants from the system-level record, and not a complete separate record.

The DOBIS access point files include name and subject authority files. The name file is both the name index to the bibliographic file and the name authority file for the system. It contains personal, corporate and conference names. Corporate and conference names are permuted to allow searching on significant keywords in the names, and a list of bilingual stopwords reduces the number of vague and general permuted entries without, it is hoped, removing any useful search terms. All name headings associated with a bibliographic record are included in the name file, regardless of whether they are main or added entries.

Authority records may be created for headings which are not yet associated with bibliographic records, as well as for headings which are associated with bibliographic records. This feature permits the addition to the file of Canadian headings established by the National Library for use by the Library of Congress, before a bibliographic record using that heading has been created. As well, authority records derived from other national MARC tapes or from other Canadian sources, for example CIP agent libraries, can also be added.

The system subject file contains headings from the most commonly used subject heading sources in Canada, such as the Library of Congress Subject Headings, Canadian Subject Headings and the Répertoire de vedettes-matière. This file is also both an index to the bibliographic file and an authority file. The subject headings are identified by type, according to recognized sources (LC, Laval) and interfiled in the same file. Cross-references refer only within one type, for instance from a variant LC heading to an acceptedc LC heading, in order to avoid potential clashes. Locally-devised subject headings are usually limited to the local file, and not recorded in the system file. As with the name file, the system and local record are linked by the same DOBIS document number, and a viewer can quickly switch between the system and local record displays. The subject file is not permuted because a Boolean search feature will be added to the system.

In creating bibliographic records on DOBIS, name, title and subject "authority" records for all required headings are added to the access point files if they are not already recorded there. The cataloguer may enter just the headings which will be sufficient to index and display that bibliographic record, or he may create a fuller authority record which will provide

cross-references, history notes, etc. for the indexes; and support the output of authority products as distinct from bibliographic products.

Authority or access point data entered on-line can be input at various levels and will be assigned a status appropriate to the source of input and the function being performed. For example, "unverified" status will be assigned to machine-readable accession reports received for union catalogue purposes from various libraries. Full catalogue records, created according to the national standard, will have a "verified" status. The various verification levels are ranked in hierarchical order from "preliminary", an acquisition record for instance, to "verified", and these levels serve to determine the bumping order for headings. The levels protect against improper changes being made on-line: cataloguers with a given security level are permitted to perform only certain functions, determined in part by the verification level of the data. The levels can also be used to determine the eligibility of data for output products - only headings with a certain status can be selected for a given product, for example, "verified" only for the production of Canadiana.

5. Authorities: National and International Networks, Problems and Benefits

The real benefits of automated authority files cannot be fully realized in batch systems, particularly in the area of networking, as the time required to integrate a new authority heading into the data base is too long. However, in an on-line environment, a new heading becomes immediately available to all users of the system whether located in the same institution or in libraries widely separated by distance. The interactive mode also enables centres of authentication to verify quickly and put their seal of approval on headings created by various members of a shared system. Both original input and authentication operations can be decentralized.

Authority file controls should be built into a system from the beginning since it is very difficult to attempt to control a data base after the fact when duplication and inaccurate matching has occurred. Despite additional development costs, the use of authority files will soon pay off. For example, in a cooperative retrospective conversion project, many headings occur over and over again, and authority files will considerably cut down on input of retrospective records as well as maintain the integrity of the data base.

In the future, the National Library of Canada intends to use on-line authority files to control a partially decentralized national bibliographic data base and network. This involves the problem of integrating records from several sources, both national and international, and of searching a variety of systems and data bases. DOBIS will be used by a network of federal government libraries, most of which have strong subject specialties, thus building a unique and valuable data base. It is likely that we will extend Cataloguing-in-Publication in a decentralized mode to federal government departments, as well as to provincial government agencies, assigning the

responsibility to departmental libraries to input bibliographic records and authority records for their own departments.

Thus, authority files are already beginning to play a part in national library systems and networks, especially in North America. What is their role on the international scene? Authority files are a component of Universal Bibliographic Control, a complementary aspect to International Standard Bibliographic Descriptions (ISBD). IFLA has already struck a Working Group on an International Authority System to recommend on an international authority system, a MARC authority format, and the exchange of data.

Authority records are not as easy to standardize as bibliographic descriptions. Headings are closely related to cataloguing rules for form and choice of name, to language, to national usages for names and to national interests for subjects. In addition it would be a great deal more difficult to divide up the responsibility for the establishment of authority records for names among countries, than it is to make each country responsible for bibliographic descriptions of their national imprint. Books are more or less clearly published within national boundaries, but people constantly move, and many change their nationality. An author's works are published in many countries. It would therefore be very difficult to avoid duplication of authority records for personal names, even if national bibliographic agencies were made responsible for establishing records for the names of their nationals who become authors. For example, a Canadian author may be first published abroad, or perhaps two countries claim the same author as their national. The situation is clearer for corporate names, and as these are much more difficult to establish than personal names, possibly this is an area on which international cooperation should concentrate. However, even corporations are multinationals!

A possible solution would be to have each country produce authority records for headings which occur in records appearing in their national bibliographies. This would result in a great deal of duplication overall, but would also provide choice in headings established according to various cataloguing rules and in various languages. These could then be exchanged on tapes or on-line from an international MARC centre. A UNIMARC authority format will need to be developed to facilitate such exchange.

Another alternative would be the development of an international standard authority number (ISAN). This would be a very complex undertaking, and I believe that such further standardization and control at an international level needs to be examined carefully. Conforming to international bibliographic standards may add an overhead to the cost of production of bibliographic data, beyond that already required by national standards. In automated systems, every change in a standard requires costly changes to complex computer systems and major updating of data bases to keep them consistent. The cost of conforming to international standards and contributing to international systems and data bases is largely borne by national bibliographic agencies, and this uses resources which such agencies could devote to other national programs. Therefore, international standardization must be of a

type that will be beneficial at the national level, for that is where the libraries are. It must also be beneficial for the end user, the researcher, and not just for the librarian middleman. The UBC programs to date, such as ISBD, do assist greatly in making bibliographic records exchangeable and usable on a world-wide basis. However, it will not be as easy to make authority records universally applicable, usable and beneficial on an international level.

Bilateral or multilateral agreements between national bibliographic agencies having the same language or using the same cataloguing rules or other standards, such as lists of subject headings, can be very beneficial. Canada and the U.S. have a number of these agreements and bibliographic cooperation between the ABACUS groups of national libraries is proving worthwhile. Cataloguing practices between Canadian and American libraries are so similar that international cooperation and standardization are clearly helpful, and this has been reflected in the name agreement between LC and NLC and in the CONSER project. However, it should be noted that both countries use the same cataloguing rules, as do the ABACUS countries, as well as the same language. As Canada is a bilingual country, English language cooperation only solves part of our cataloguing problem.

It will be essential to find the proper level of international cooperation on authority files. One possibility is a basic level which establishes names and variants, especially for corporate bodies, as neutrally and independent of cataloguing rules as possible, coding all the name elements so that they can be manipulated for the required usage of particular cataloguing rules, for example, as parts and relationships of a hierarchy. Higher levels of international exchange of authority records could take place among countries having common systems, data bases, languages, cataloguing standards, scripts, etc. The cost of encoding elements in authority files beyond that needed to maintain and control national bibliographic data bases must be carefully considered in relation to concrete benefits received. I believe that authority files are necessary on a national level to support a national bibliographic data base whether centralized or decentralized. As well, some exchange of authority data on an international level will be useful, particularly for multilingual data bases. The Canadian system has shown that language-linked equivalences work well in a bilingual country, experience will show how many languages can be effectively and economically linked.

In closing, I would like to summarize the benefits to be gained from an automated authority file system.

First and foremost are data base and file integrity. Each heading occurs only once in the data base and is a unique heading. This ensures that all works related to a name will be properly collocated under that name and enables a user to make a comprehensive search. This heading integrity also helps to eliminate record duplication in the data base, reduces the number of errors due to incorrect copying or keying of data, and facilitates the matching of records coming from many sources.

The ability to rapidly update the access points and ensure that the most current form of name is consistently used in all records in which it occurs enables name changes, changes in cataloguing rules and in subject terminology to be economically carried out. The most up-to-date headings are used in all bibliographic products created from the data base.

Since an authority heading occurs only once in the data base, great savings are achieved in time and money in coding of bibliographic records, data input, proofreading and verification and in data storage, thus eliminating data redundancy and reducing the size of the data base.

An authority file also serves to customize records and products for individual network members, by providing flexibility in choice of headings for local usage, by making it possible to relate local records to the network data base, and by automatically substituting access points in another language through the use of linked equivalents. Authority files can thus be used to "nationalize" records received from other countries, by substituting headings constructed according to the national cataloguing and subject standards in the national language or languages. In the same manner, records can be upgraded from local to system standards by matching against a system's authority file. This will assist decentralized shared cataloguing and retrospective conversion projects while maintaining file integrity.

Authority files are the way of the future. They allow local or national flexibility while ensuring network compatibility, and make possible the national and international exchange and exploitation of bibliographic data on a scale never before possible. In the future, they will play a role in bringing into full implementation the international program of UBC, and as a result, that of UAP, making the world's literature available to all who need it.

Acknowledgement

I wish to express my appreciation of the considerable contribution of my Special Projects Officer, Mrs. Elizabeth Hawkins, in the preparation of this paper.

Bibliography

Buchinski, Edwin, J., William L. Newman and Mary Joan Dunn. "The automated authority subsystem at the National Library of Canada." Journal of Library Automation. 9(4), Dec. 1976. p. 279-298.

Buchinski, Edwin, J., William L. Newman and Mary Joan Dunn. "The National Library of Canada authority subsystem: implications." Journal of Library Automation. 10(1), Mar. 1977. p. 28-40.

Dunn, Mary Joan. "Automation at the National Library: The Canadiana/Cataloguing Subsystem". Canadian Library Journal. 33(4), Oct. 1976.

Forget, Louis J.S. "Automation at the National Library of Canada". Paper presented at the IFLA World Congress of Librarians, Library Automation Seminar. La Hulpe, Belgium, Sept. 6-7, 1977. 50 p.

What's in a name?: control of cataloque records through automated authority files. Edited and compiled by Natsuko Y. Furuya. Toronto, UTLAS, 1978. 230 p. Proceedings of the workshops sponsored by National Library of Canada, Canadian Library Association, Office of Library Coordination, Council of Ontario Universities, University of Toronto Library Automation Systems.

Exhibits

ure 1

CORD CREATION PROCESSES

ional Library of Canada

omated Systems

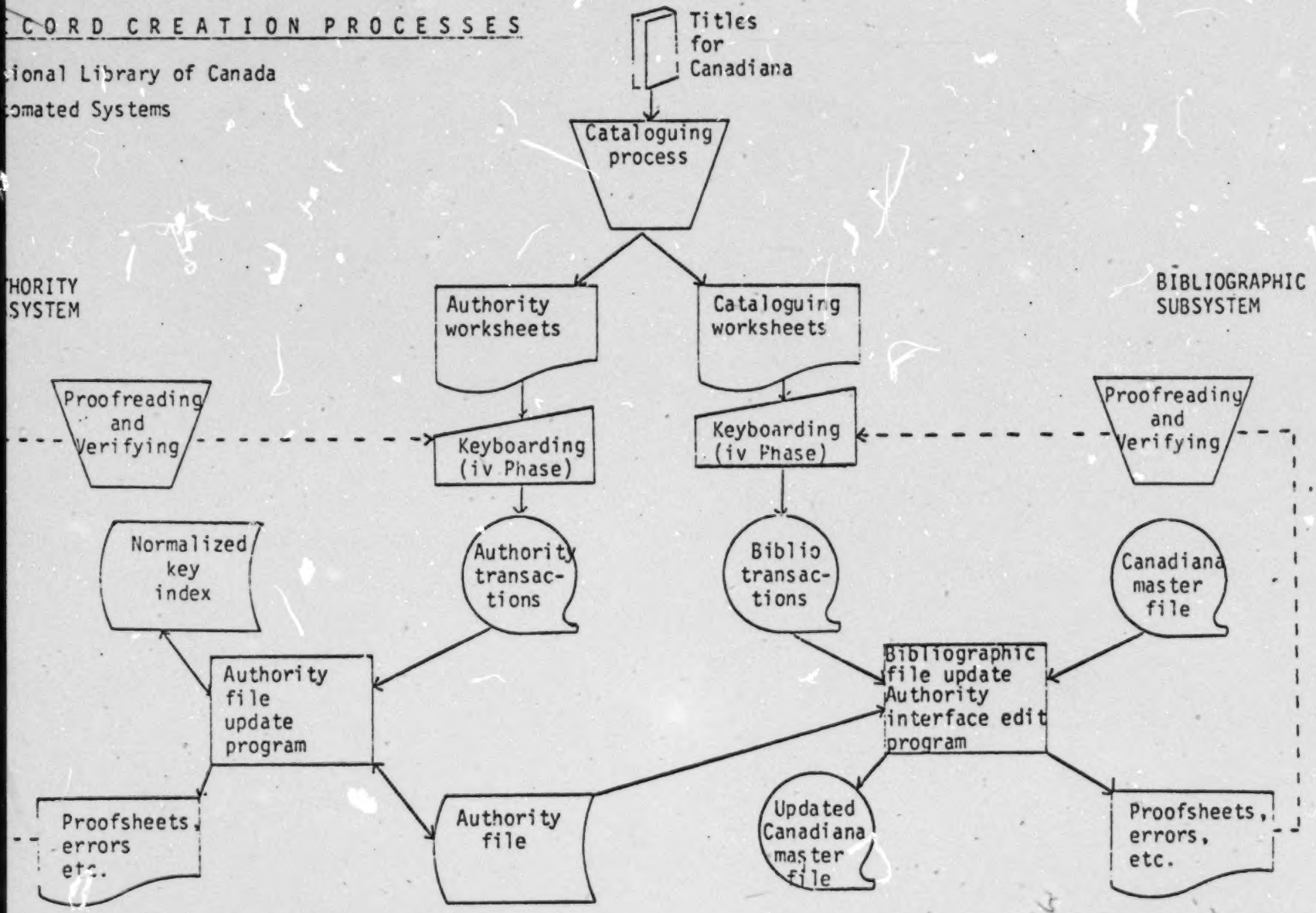
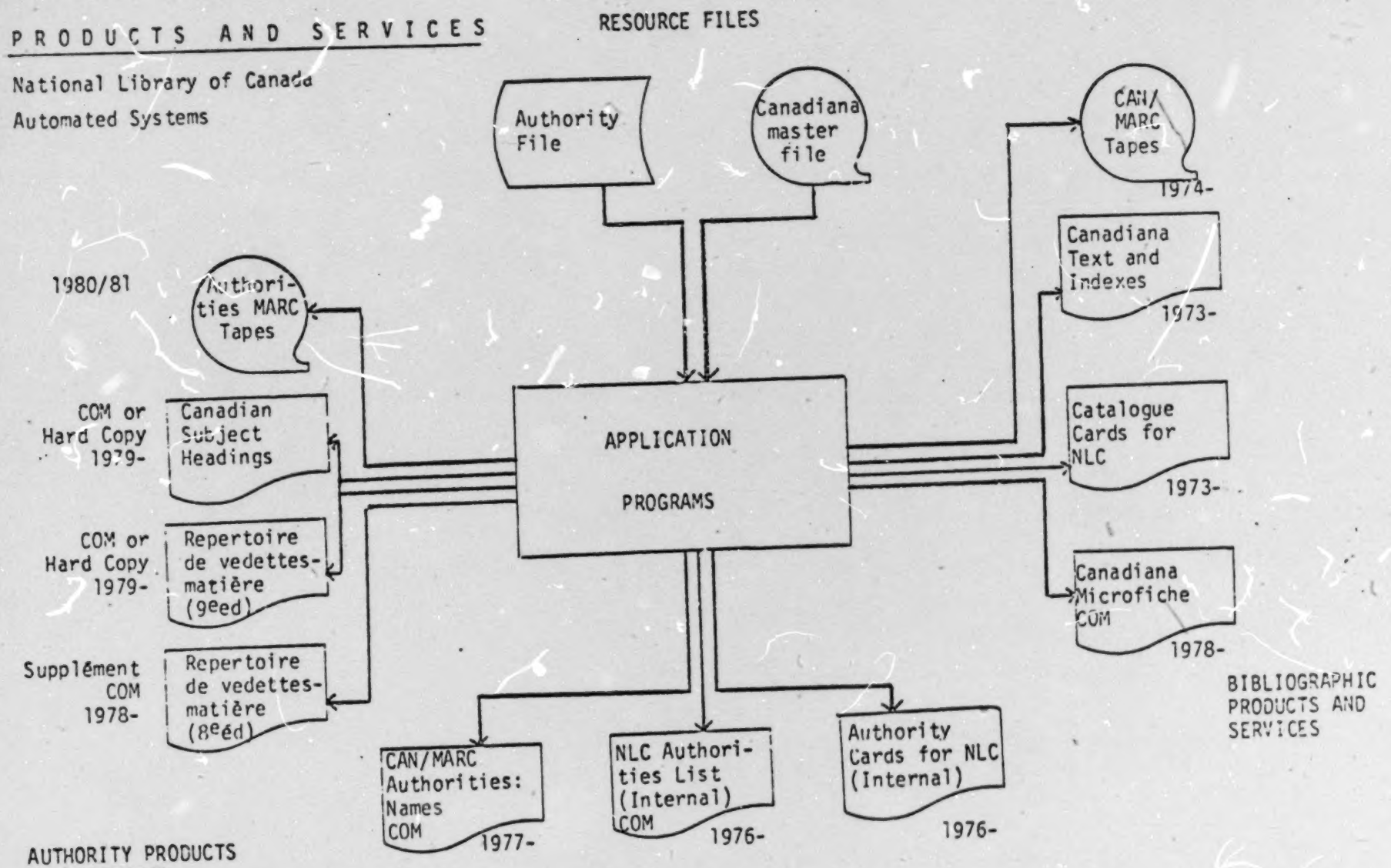
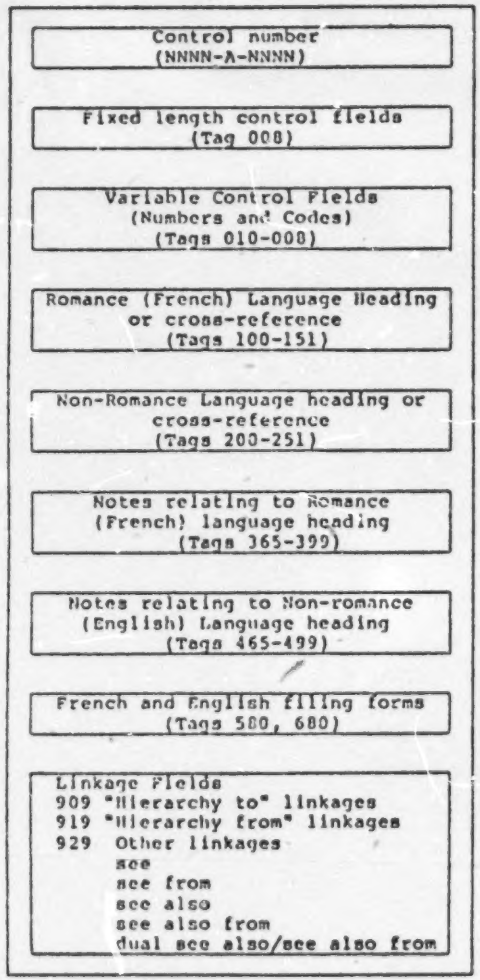


Figure 2


PRODUCTS AND SERVICES
RESOURCE FILES
National Library of Canada
Automated Systems
Authority File
Canadiana master file
CAN/MARC Tapes
1974-
Canadiana Text and Indexes
1973-
1980/81
Authorities MARC Tapes
COM or Hard Copy 1979-
Canadian Subject Headings
APPLICATION PROGRAMS
Catalogue Cards for NLC
1973-
COM or Hard Copy 1979-
Repertoire de vedettes-matière (9eed)
Canadiana Microfiche COM
1978-
Supplément COM 1978-
Repertoire de vedettes-matière (8eéd)
BIBLIOGRAPHIC PRODUCTS AND SERVICES
CAN/MARC Authorities: Names COM
1977-
NLC Authorities List (Internal) COM
1976-
Authority Cards for NLC (Internal)
1976-
AUTHORITY PRODUCTS AND SERVICES

NATIONAL LIBRARY INTERNAL AUTHORITY RECORD FORMATS



SCHEMATIC RECORD FORMAT

Record for Heading

0001-K-1089

110 Commission canadienne du lait.

210 Canadian Dairy Commission.

365 Société de la couronne établie le 11 juillet 1966.

465 Crown corporation established July 11, 1966.

929 0001-F-1090 (see from)

929 0001-E-1091 (see from)

Separate Records for Cross References

0001-F-1090

110 Canada. Commission canadienne du lait.

210 Canada. Dairy Commission.

929 0001-K-1089 (see)

0001-E-1091

110 Canada. Ministère de l'agriculture. Commission canadienne du lait.

210 Canada. Dept. of Agriculture. Dairy Commission.

929 0001-K-1089 (see)

Figure 4

LINKAGES BETWEEN AUTHORITY RECORDS IN NLC AUTHORITY SUBSYSTEM

These schematic records in the NLC internal format illustrate a multiple change in which the English and French forms not only change, over time, but independently in the two languages. To simplify the exhibits, other linkages between these records and other records in the NLC system are not shown (e.g. "hierarchy to" and "hierarchy from" references).

0000-H-0046

110 Canada. Ministère des transports

210 Canada. Dept. of Transport

365 Etabli le 2 nov. 1936, par la fusion du Ministère de la marine et le Ministère des chemins de fer et canaux.

465 Established Nov. 2 1936, absorbing the former Dept. of Marine and Dept. of Railways and Canals. Name changed Feb. 19, 1970 from Dept. of Transport to Ministry of Transport.

680 Canada. Department of Transport.

929 Linkages:
0000-H-0038 (earlier name, English and French)
0000-G-0039 (earlier name, English and French)
0000-G-0047 (French equivalence)
0000-G-0047 (later name in English)
0000-F-0048 (see from reference, English and French)

Record for Current Name

0000-G-0047

110 0000-H-0046 (French form of name stored in the other record)

210 Canada. Ministry of Transport

465 Name changed Feb. 19, 1970, from Dept. of Transport to Ministry of Transport.

929 Linkages:
0000-H-0046 (French equivalence)
0000-H-0046 (earlier name in English)
0000-F-0048 (see from reference, English and French)

Record for Current Name

0000-H-0038

110 Canada. Ministère des chemins de fer et canaux

210 Canada. Dept. of Railways and Canals

365 Etabli le 13 mars 1879. Fusionné avec le Ministère de la marine le 2 nov. 1936 pour former le Ministère des transports.

465 Established Mar. 19, 1879. Merged with the Dept. of Marine on Nov. 2, 1936 to form the Dept. of Transport.

680 Canada. Department of Railways and Canals.

929 Linkages:
0000-H-0046 (later name, English and French)

Record for Earlier Name

0000-F-0048

110 Transports Canada

210 Transport Canada

929 Linkages:
0000-H-0046 (see reference, English and French)
0000-H-0047 (see reference, English and French)

Record for Cross Reference

0000-G-0039

110 Canada. Ministère de la marine

210 Canada. Dept. of Marine

365 En 19307 le Ministère de la marine et des pêcheries a divisé pour former le Ministère de la marine et le Ministère des pêcheries. Le Ministère de la marine a fusionné avec le Ministère des chemins de fer et canaux le 2 nov. 1936 pour former le Ministère des transports.

465 In 19307 the Dept. of Marine and Fisheries split into the Dept. of Marine and the Dept. of Fisheries. The Dept. of Marine merged with the Dept. of Railways and Canals on Nov. 2, 1936 to form the Dept. of Transport.

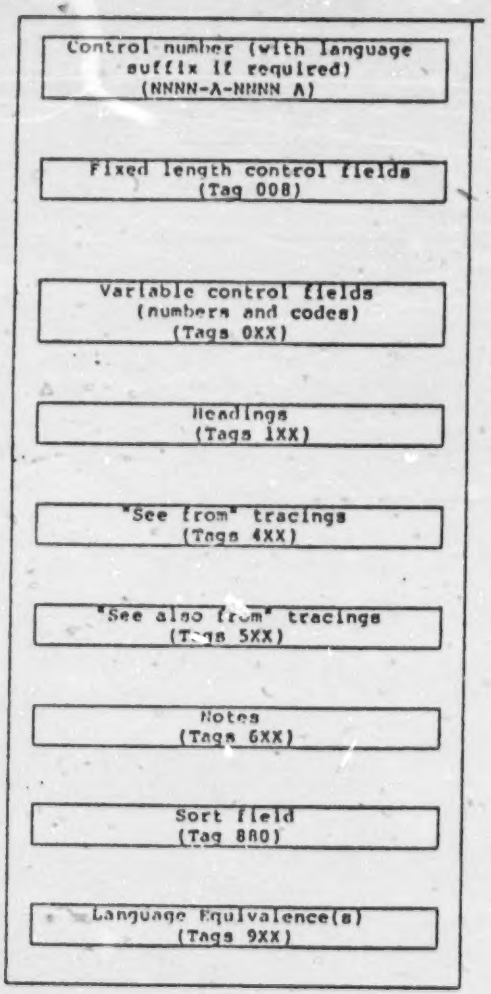
680 Canada. Department of Marine

929 Linkages:
0000-G-0040 (earlier name, English and French (record not shown here))
0000-H-0046 (later name, English and French)

Record for Earlier Name

ure 5

CANADIAN MARC COMMUNICATION FORMAT: AUTHORITIES


Control number (with language suffix if required) (NNNN-A-NNNN A)
Fixed length control fields (Tag 008)
Variable control fields (numbers and codes) (Tags 0XX)
Headings (Tags 1XX)
"See from" tracings (Tags 4XX)
"See also from" tracings (Tags 5XX)
Notes (Tags 6XX)
Sort field (Tag 880)
Language Equivalence(s) (Tags 9XX)

SCHEMATIC RECORD FORMAT

0001-R-1089E

110 Canadian Dairy Commission.

410 Canada. Dairy Commission.

410 Canada. Dept. of Agriculture. Dairy Commission.

665 Crown corporation established July 11, 1966.

910 Commission canadienne du lait

English Authority Record

0001-R-1089F

110 Commission canadienne du lait.

410 Canada. Commission canadienne du lait

410 Canada. Ministère de l'agriculture. Commission canadienne du lait

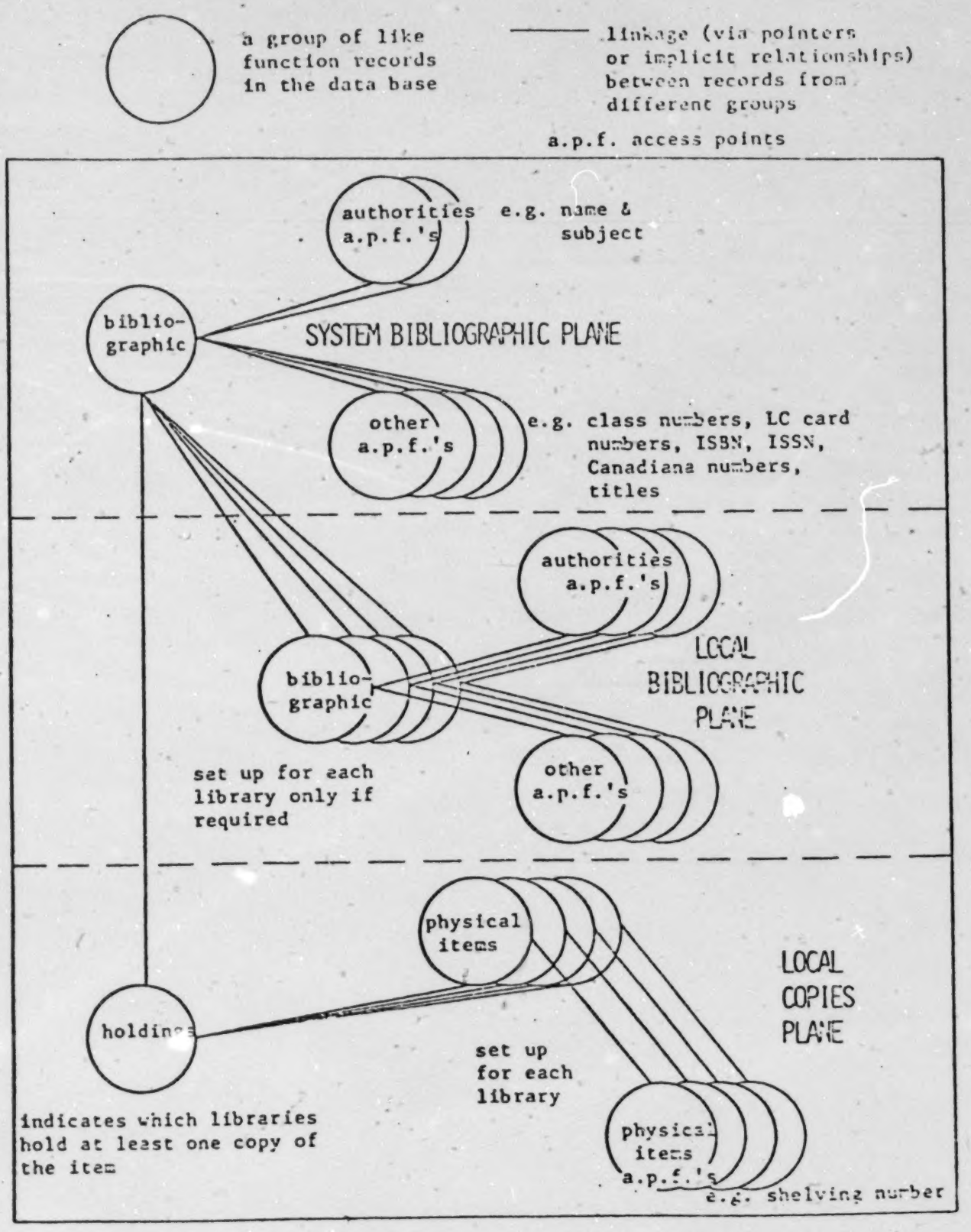
665 Société de la couronne établie le 11 juillet 1966.

910 Canadian Dairy Commission.

French Authority Record

SAMPLE AUTHORITY RECORDS

DOBIS DATA BASE STRUCTURE OVERVIEW


a group of like function records in the data base
linkage (via pointers or implicit relationships) between records from different groups
a.p.f. access points
authorities a.p.f.'s
e.g. name & subject
bibliographic
SYSTEM BIBLIOGRAPHIC PLANE
other a.p.f.'s
e.g. class numbers, LC card numbers, ISBN, ISSN, Canadiana numbers, titles
authorities a.p.f.'s
bibliographic
LOCAL BIBLIOGRAPHIC PLANE
other a.p.f.'s
set up for each library only if required
physical items
LOCAL COPIES PLANE
holdings
set up for each library
indicates which libraries hold at least one copy of the item
physical items a.p.f.'s
e.g. shelving number